NIH〉 Center for Scientific Review
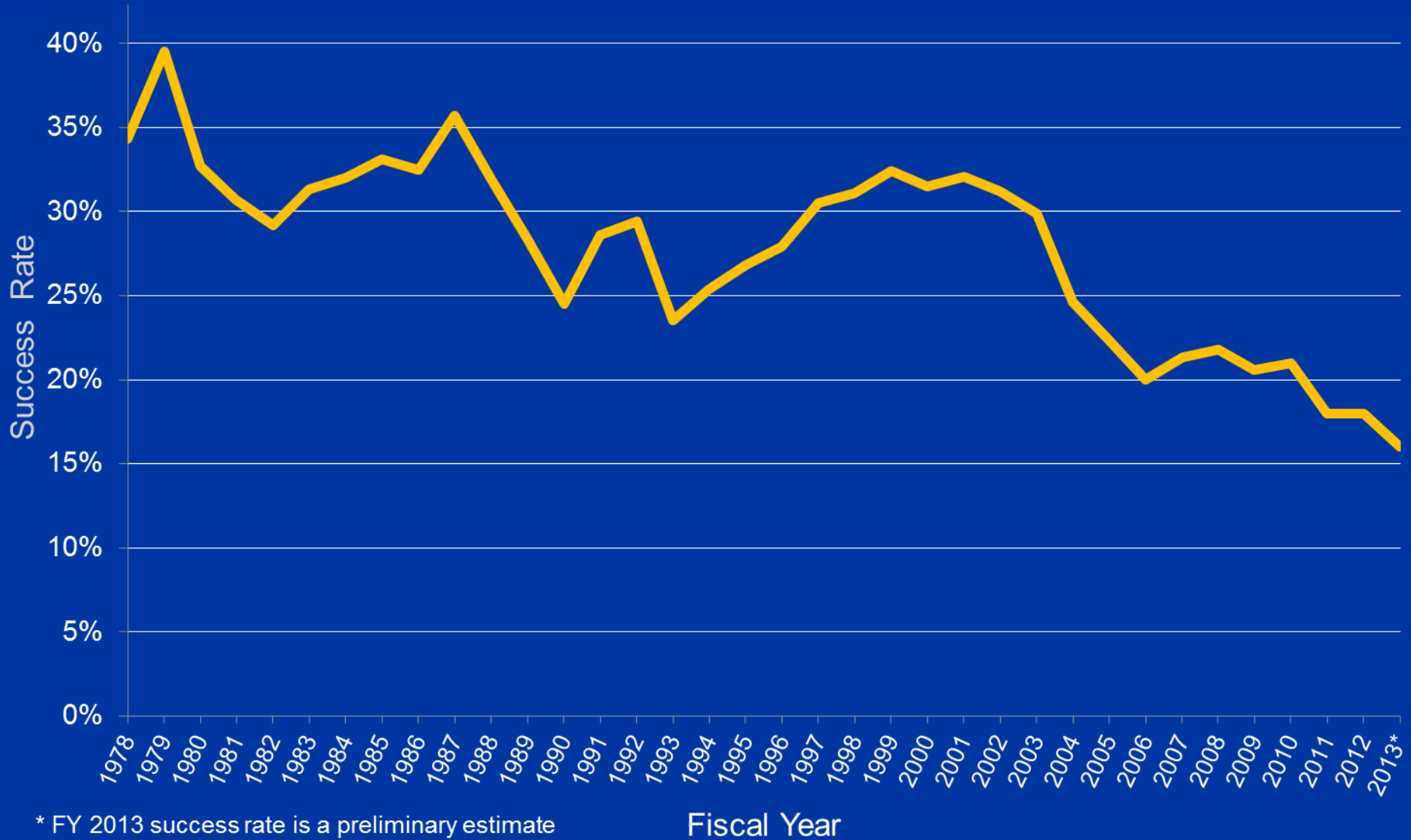
Measuring Quality in CSR

Richard Nakamura, Ph.D.

December 16, 2013

Grant Success Rate
1978 - 2013

* FY 2013 success rate is a preliminary estimate

CSR Mission

To see that NIH grant applications receive fair, independent, expert, and timely reviews – free from inappropriate influences – so NIH can fund the most promising research.

Possible measures related to quality in peer review

CSR encourages an approach to measuring quality with the following principles:

- Is based in scientific judgment
- Is validated via feedback based on actual scientific outcomes
- Is tested to ensure that any quality measure delivers benefit and not harm
- Is systematically compared against a broad set of putative quality measures

I. Use of scientific judgment (Peer review) – CSR

a) The Standard has been peer review of applications with standing study sections, strong member reviewers in a face to face format. Can we do better?

b) Comparing top papers emerging from past awards using science judges (long feedback loop and labor intensive but most confidence in validity). DPCPSI contract data

c) Compare outputs of individual SRGs by reranking best scored applications across study sections within IRGs. Improves speed of feedback. Preliminary results available

d) Examining the reliability of scoring by reviewing the same application by more than one SRG. This will provide basic statistics that we need for reliability and power calculations before validity studies can be done. In planning

II. Evaluation of the distribution of applications

Our system of percentiling assumes that the highest quality applications are distributed evenly across study sections. Is this true?

Preliminary results available

III. Surveys of scientists and science administrators – CSR

Collecting comparative opinions of reviewers and applicants. What do scientists want for themselves?

    a. Comparing review platforms

    b. Comparing committees

    c. Comparing SROs, Program staff

Survey created, implementation held up by shutdown

IV. IMPAC database analyses - OER and CSR

a. Success rates of PIs, departments, institutions, SRGS, IRGs

b. Application rates

c. Rates of type 2 application

Some results available in ARGO from OER

V. DRR Database - CSR

a. Requests for assignment

b. Alternate choices and negatives

c. Moving from prescriptive to investigator choices- considers demand

OPAE has been developing analyses and with DRR some tools

VI. Text analysis – OER, CSR

a. Compare success of conforming and nonconforming applications (like functions)   Preliminary studies conducted by Sy Garte of CSR

b. Text analysis of applications and critiques for grammar and vocabulary – OER, CSR Planned and funded application

c. Text analysis for evaluation of summary statement standards

d. Examine applications for changing ideas and rapidity of change

VII. Bibliometric – DPCPSI, CSR and others

a. Citations attributable to grants post award – long feedback loop Results from DPCPSI

b. Citation analysis of reviewers and applicants may provide information about how science is organized.

c. Out-performing journal impact factors – Relative Citation Rating – Results from DPCPSI

d. Examine networks of PIs and reviewers with co-citations and collaborations . Preliminary results from DPCPSI

e. Split awarded applications in the top 10 and the next 10 percentile (from 10 years ago) to see if there is a difference in citation activity – this is about the ability of review to differentiate.

VIII. Network or structural analysis based on publications or co-citations – CSR

a. How well is the science covered by CSR IRGs and SRGs?
   Mappings of science for CSR shown previously
b. Measuring currency of ideas in SRGs
c. Importance of contribution to developing areas of biomedical science
d. Consider trend analysis
   •

IX. Other- Possible for studies from outside of NIH collaborations wanted

a. Understanding sources of recognized major advances (prizes)
b. Doing a factor analysis of quality measures – relationship to long term success
c. Factor analysis of high performing awards
d. Compare rankings and performance of cross organizational / cross national funding groups
e. Historical analysis of scientific advances
f. Econometric studies
g. Consider prizes for systematic approaches to understanding great advances in science and medicine
   i. How do we determine the history of advances?
   ii. How can we maximize the yield of science investment?

- Finally do analyses of the performance of quantitative measures against the long range analyses of award productivity

Two examples

Intra IRG - Cross SRG Application Ranking Study

Is the quality of highly ranked applications the same across study section?

IRG Participants

- GGG - Genes Genomes and Genetics - Panniers
- PSE - Populations Sciences and Epidemiology - Durrant
- BDCN - Brain Disorders and Clinical Neuroscience - Edwards
- HDM - Healthcare Delivery and Methodologies- Fosu
- IDM - Infectious Diseases and Microbiology - Politis

The CSR Ranking Pilot

- Evaluating the quality of applications across study sections
- 5 Integrated Review Groups
- X study sections
- N reviewers
- P applications

- α 0.05
- power 0.85-0.9
- non-parametric ANOVA for analysis

CSR Ranking Pilot
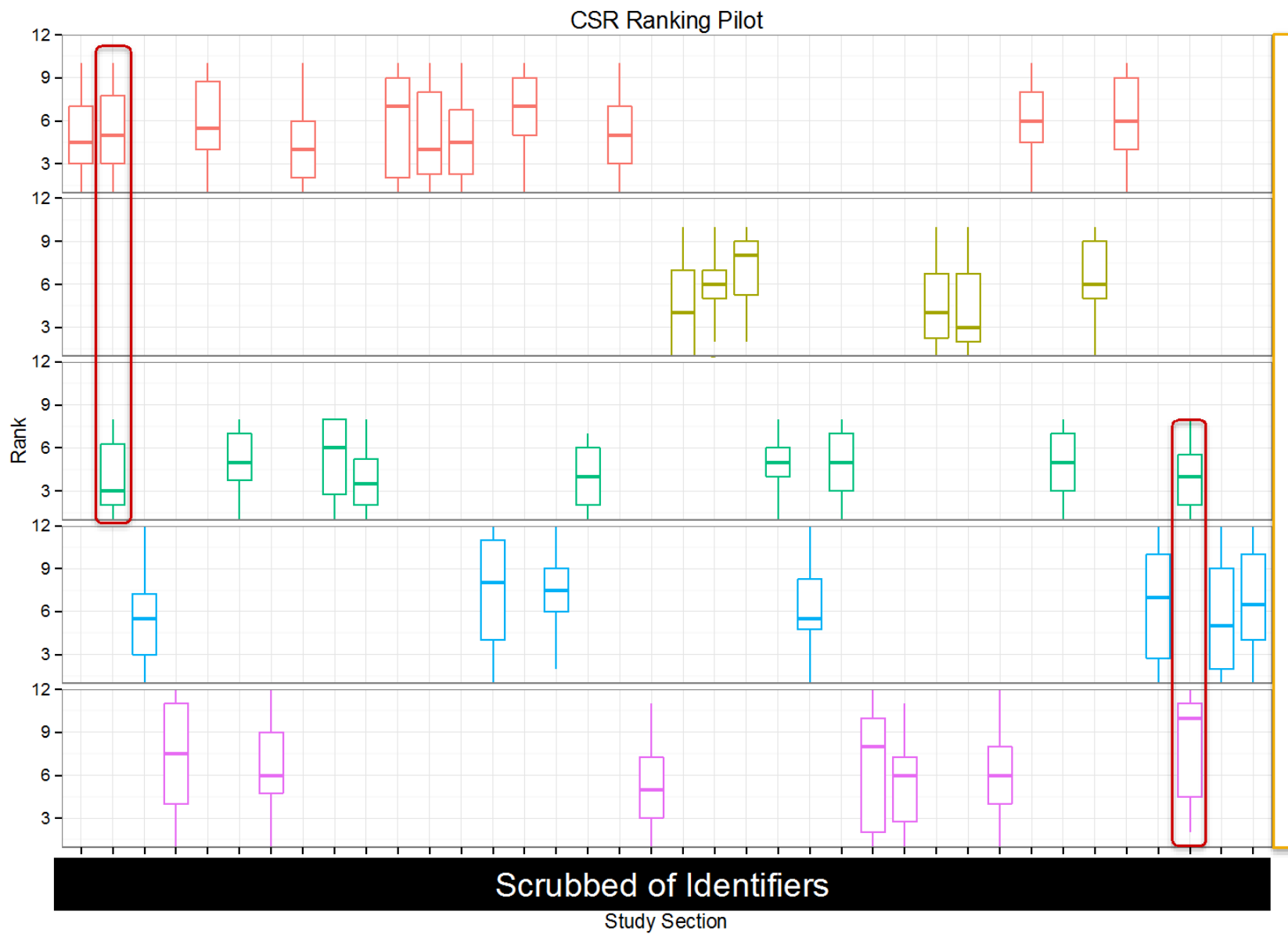
Genes, Genomes and Genetics (GGG) IRG

- GCAT Genomics, Computational Biology and Technology

- MGA Molecular Genetics A

- MGB Molecular Genetics B

- GHD Genetics of Health and Disease

- PCMB Prokaryotic Cell and Molecular Biology

- GVE Genetic Variation and Evolution

The CSR Ranking Pilot
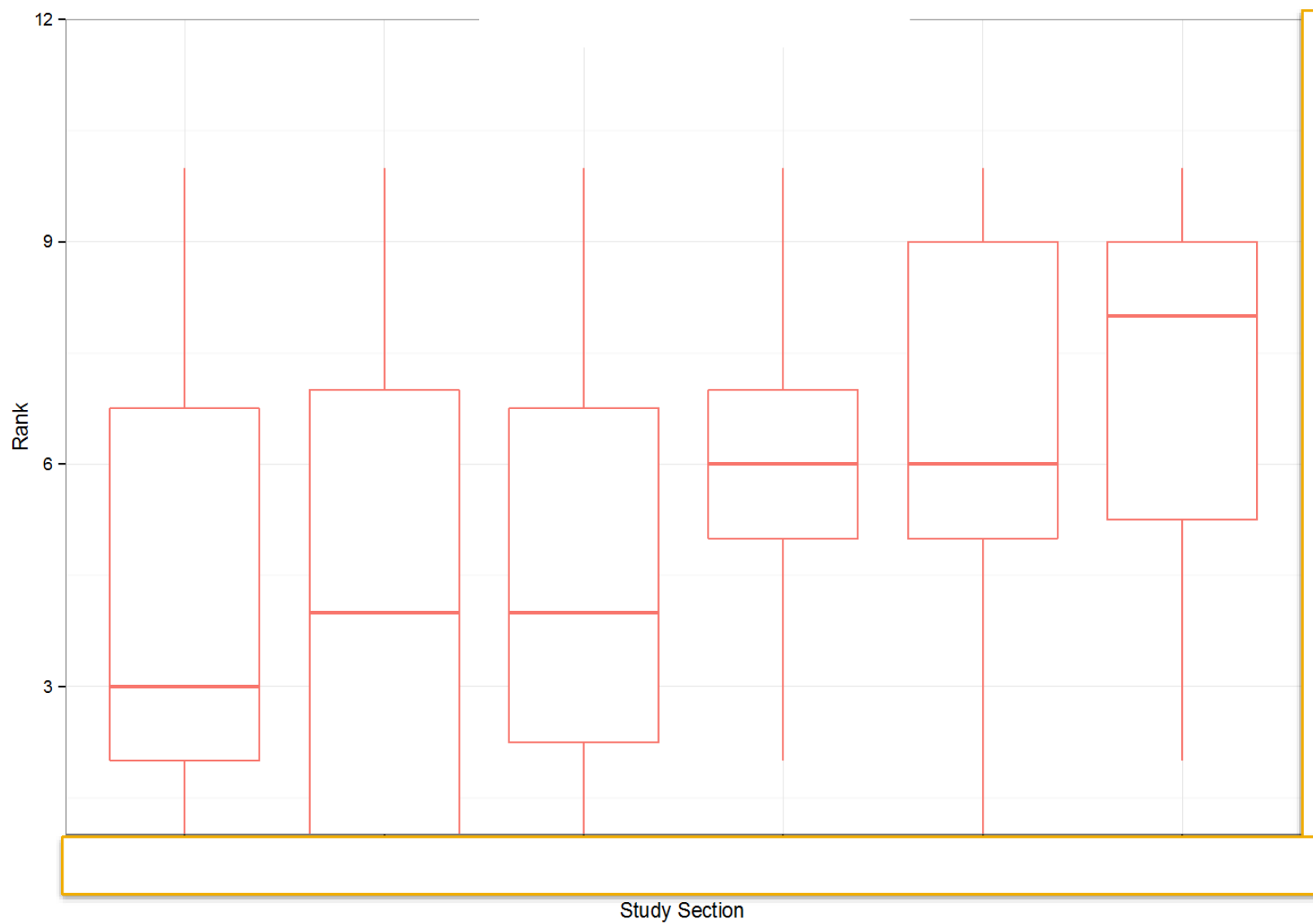
Tolerance for innovation

The Issue

- How well do study section members react to innovative science outside the standard concepts and approaches in their fields?

- The question of conformity vs. innovation in NIH peer review has been raised in the scientific and popular press.

Methodological Approach

- Compute a QVR fingerprint in for all applications submitted to each study section from 2007 to 2011. (146 SRGs)

- Using QVR LIKE function, determine the Match Score (MS) for each application.

- Applications with higher match scores are more conforming to the standard applications

Negative Correlation between SII and Bibliometric Performance (p < 0.001)
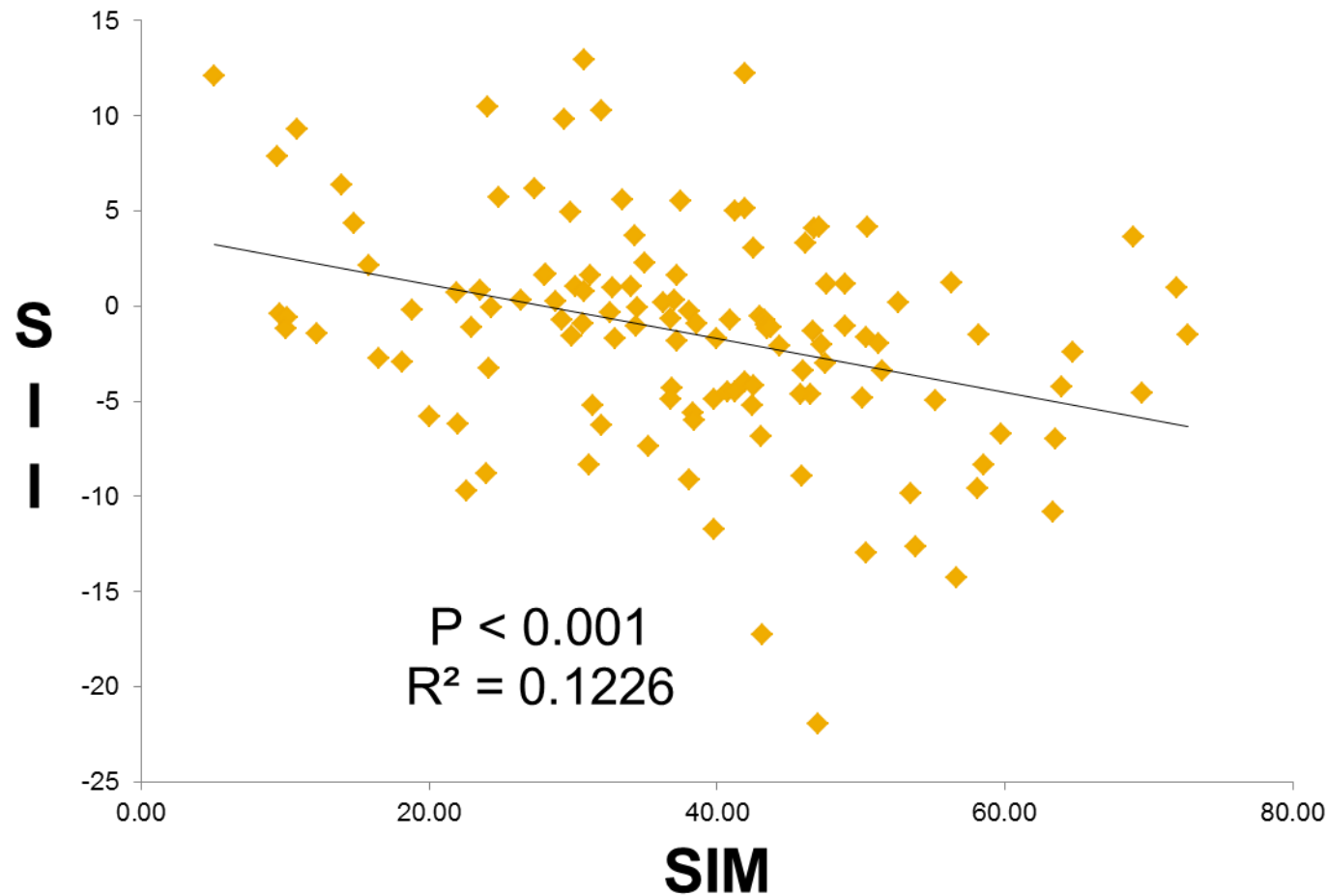
Summary and Conclusions

- The great majority (90%) of CSR Study Sections do not penalize applications that do not conform

- The basic science SRGs tend to be more conformist that translational or clinical SRGs

- There is a wide distribution among IRGs, with some SRGs showing different behavior with any IRG

- There is an Inverse correlation between innovation and bibliometric measure of quality.

In our studies the goal is to produce publications subject to peer review. So there may be a significant lag before policy change.

Questions? Comments?

CSRDirector@csr.nih.gov